Healthcare Twitter Analysis

I am new to this entire subject area

- Twitter data
- Data Mining
- R, Python
- Medical research

... so I will make no claim to provide any breakthroughs but I intend to proceed as follows, following Pratik's framework:

Algorithms Research

I have already started writing programs to pull in additional data to supplement the files provided. I have also begun statistical and textual analyses on the Twitter data. See GitHub.

Technology / Framework Research

It is conceivable that this mass of data will turn out to be too big for my little laptop. I will look into setting up an environment on AWS. Even if it isn't strictly necessary, I am interested in this and will attempt to set it up just for my own education. Who knows ... Mongo DB, Hadoop? We'll see.

Business/ Domain Research

It seems clear to me that the Twitter data itself is of little use for the aims of this project so I will look for other data sources that will usefully supplement it. My domain knowledge is essentially nil, so without collaborative help I am unlikely to break new ground.

Visualization Research

Aside from what naturally flows from the data & statistical analyses, I am somewhat suspicious of this area ... my observation is that it is all-too-easy to create very beautiful pictures from vacuous underlying analyses; too many people start here and ultimately fail to add any value. If I get to this point at all, I will attempt to add value with pictures rather than creating them for their own sake.

Collaboration

My offer of collaboration on the discussion board went unanswered, but let me say here I would be pleased to work with other people. In any event, I intend to post all of my work on GitHub for others to view and use.

George Fisher
Boston July 2014